

# Identification of circles from datapoints using Gaussian sums

T. Alexopoulos<sup>1</sup>, G. Iakovidis<sup>1,2</sup>, S. Leontsinis<sup>1,2\*</sup>, K. Ntekas<sup>1,2</sup> and V. Polychronakos<sup>2</sup>

<sup>1</sup>*National Technical University of Athens*

<sup>2</sup>*Brookhaven National Laboratory*

March 25, 2014

## Abstract

We present a pattern recognition method which use datapoints on a plane and estimates the parameters of a circle. MC data are generated in order to test the method's efficiency over noise hits, uncertainty in the hits positions and number of datapoints. The scenario where the hits from a quadrant of the circle are missing is also considered. The method proposed is proven to be robust, accurate and very efficient.

## 1 Introduction

Various types of detectors require algorithms that handle number of datapoints in a plane and reconstruct the parameters of a circle, [1],[2],[3]. These algorithms need to be robust against noise hits and analyse big amount of data relatively fast. Finally, the performance of the algorithm needs as independent as possible with the resolution of the detector.

In a recent paper [4] we examined the use of the Legendre transform for the determination of a circle's characteristics. In this paper we revisit this problem and describe in detail an alternative method, based on Gaussian sums, proposed for the reconstruction of the circle's center  $(x_0, y_0)$  and radius  $R$  from a given set of datapoints.

## 2 Description of the Methods

Following the notation and train of thought of [4], using 3 datapoints  $(\mathbf{r}_1 = A(x_1, y_1), \mathbf{r}_2 = B(x_2, y_2)$  and  $\mathbf{r}_3 = C(x_3, y_3))$ , we can have a first estimation of the center  $(x_{\text{est}}, y_{\text{est}})$  and the radius,  $R_{\text{est}}$ , of the circle (see figure 1).

These estimates are given by:

$$\begin{aligned} x_{\text{est}} &= \frac{m_t m_r (y_3 - y_1) + m_r (x_2 + x_3) - m_t (x_1 + x_2)}{2(m_r - m_t)} \\ y_{\text{est}} &= -\frac{1}{m_r} \left( x_{\text{est}} - \frac{x_1 + x_2}{2} \right) + \frac{y_1 + y_2}{2} \\ R_{\text{est}} &= \sqrt{(x_{\text{est}} - x_1)^2 + (y_{\text{est}} - y_1)^2} \end{aligned}$$

where  $m_t = (y_3 - y_2)/(x_3 - x_2)$  and  $m_r = (y_2 - y_1)/(x_2 - x_1)$  the slopes of the lines connecting  $AB$  and  $BC$ .

---

\*stefanos.leontsinis@cern.ch

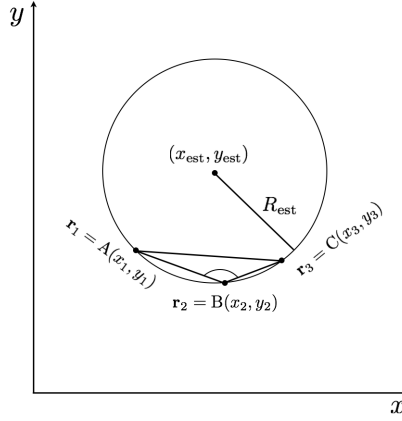


Figure 1: Representation of a circle with center  $(x_{\text{est}}, y_{\text{est}})$  and radius  $R_{\text{est}}$ .

## 2.1 Gaussian Sum

Having  $n$  given datapoints/measurements,  $k = \binom{n}{3}$  different centers and radii can be estimated. Using the formulas described above the global sums are constructed

$$G(x) = \sum_{i=1}^k \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(x-x_{\text{est}}^i)^2/2\sigma_x^2}$$

$$G(y) = \sum_{i=1}^k \frac{1}{\sqrt{2\pi}\sigma_y} e^{-(y-y_{\text{est}}^i)^2/2\sigma_y^2}$$

$$G(R) = \sum_{i=1}^k \frac{1}{\sqrt{2\pi}\sigma_R} e^{-(R-R_{\text{est}}^i)^2/2\sigma_R^2}$$

where  $x_{\text{est}}^i$ ,  $y_{\text{est}}^i$  and  $R_{\text{est}}^i$  are the circle's parameters for the  $i^{\text{th}}$  triplet of datapoints that defines the circle. The standard deviations of the gaussians are set of a fraction of the geometrical scale of the problem. In the studies below  $\sigma_x = \sigma_y = \sigma_R = 0.1$  have been used. The reason of using gaussian functions in the definition of the global sums (see above) instead of using the basic  $\chi^2$  test functions, is based on the fact that noise type datapoints will give values of the parameters far away from the right answer. These estimated parameters ( $x_{\text{est}}^i$ ,  $y_{\text{est}}^i$  and  $R_{\text{est}}^i$ ) will contribute very little to the global sums. On the contrary, the right values of the estimated parameters will give the maximum contribution.

For each event the functions described above (see Figure 2) are constructed and the value of the bin with the maximum number of entries is chosen. The value of these bins are the estimate of the  $(x_{\text{est}}, y_{\text{est}})$  and  $R_{\text{est}}$  of the circle.

## 2.2 Extraction of the Circle Parameters

The extraction of the circle parameters using the first method (see section 2.1) is straightforward: the bin with the maximum number of entries per event in each  $G(x)$ ,  $G(y)$  and  $G(R)$  gives the estimate of the center's coordinates and the radius of the circle. An example event can be seen in Figure 2. In Figure 2(a) the circle used to generate the datapoints is shown. The  $G(R)$ ,  $G(x)$  and  $G(y)$  distributions for the circle parameters can be seen in Figures 2(b), 2(c) and 2(d) respectively.  $G(x)$  and  $G(y)$  for  $x_{\text{est}}$  and  $y_{\text{est}}$  peak at 6 where the center of the circle is and  $G(R)$  peaks at  $R_{\text{est}} = 3$  indicating the actual radius of the circle.

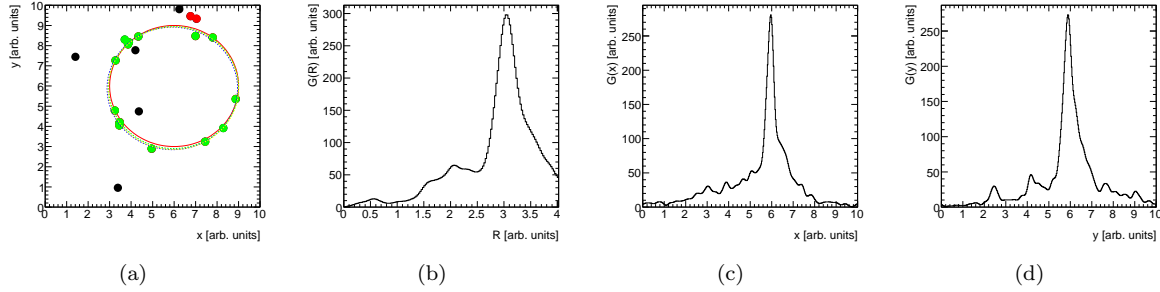


Figure 2: (a) The red datapoints originate from the circle (red line) having received a smearing of 10%. The datapoints that missed the radius  $R$  (noise hits - black datapoints) are on a 50% percentage of the circle's datapoints. Red circle is the original circle used to generate the datapoints, blue is reconstructed by using  $\chi^2$  minimization, green is reconstructed by using the Legendre technique and yellow using the gaussian sum. (b) Example of the  $G(R)$  function for the reconstruction of the radius  $R$ . (c) Example of the  $G(x)$  function for the reconstruction of the  $x_0$ . (d) Example of the  $G(y)$  function for the reconstruction of the  $y_0$ .

### 3 Performance Studies

The algorithm is tested using simulated experiments, creating datapoints ( $n = 8 - 20$ ) that belong to a circle. The robustness of the algorithms was tested by inserting gaussian uncertainties in both the  $x$  and  $y$  coordinates of the datapoints/measurements and random noise hits/outliers.

A special scenario was considered, where one quadrant of the detector was deactivated (results are summarized in table 1).

#### 3.1 One Circle Case

The center of the circle was chosen to be at  $(4, 5)$  with a radius of 2 (arbitrary units). Using this circle,  $n$  datapoints ( $n = 8, 10, 15$  and  $20$ ) are randomly created. We smear both the  $x_i$  and  $y_i$  positions of the datapoints, independently on  $x$  and  $y$ , by 5%, 10% or 15% of the circle's radius  $R$ . Apart from the uncertainty introduced in the position of the datapoints, noise hits in certain percentage of the datapoints that originate from the circle are included. The noise hits are randomly generated in the ranges  $0 \leq x \leq 10$  and  $0 \leq y \leq 10$ . An event generated by this toy MC can be seen in Figure 2(a). The red and green datapoints ( $n = 15$ ) are generated from the circle and the black datapoints are the noise hits (50% noise level is used in this event). The positions of all the generated datapoints  $(x_i, y_i)$  are smeared by 10% of its radius  $R$ .

The results of the methods described in section 2 are summarized in Table 2. The algorithm proposed for the extraction of the circle's parameters work well under all test scenarios.

Table 1: Results of the algorithms described in section 2.2. We use  $n = 10$  datapoints, generated in three quadrants of the circle, three different percentages of position uncertainty (5%, 10% and 15% of the radius  $R$ ) and five additional noise hits. The original values of the circle's parameters can be found in the first row.

Original Values	$R = 3$	$x = 6$	$y = 6$
Uncertainty [%]	5	10	15
R	$3.01 \pm 0.00$	$3.03 \pm 0.00$	$3.04 \pm 0.01$
x	$6.00 \pm 0.00$	$5.99 \pm 0.01$	$6.00 \pm 0.02$
y	$6.00 \pm 0.00$	$5.99 \pm 0.01$	$5.98 \pm 0.01$

Table 2: Results of the algorithms described in section 2.2. We try four different sets of number of datapoints ( $n = 8, 10, 15, 20$ ), three different percentages of position uncertainty (5%, 10% and 15% of the radius  $R$ ) and three levels of noise hits (0%, 25% and 50% of the actual datapoints). The original values of the circle's parameters can be found in the first row.

Original Values	$R = 2$			$x = 5$			$y = 4$		
$n = 8$									
Noise [%]	0			25			50		
Uncertainty [%]	5	10	15	5	10	15	5	10	15
R	$2.00 \pm 0.00$	$2.01 \pm 0.00$	$2.02 \pm 0.01$	$2.01 \pm 0.00$	$2.02 \pm 0.00$	$2.02 \pm 0.01$	$2.01 \pm 0.00$	$2.02 \pm 0.00$	$2.04 \pm 0.01$
x	$5.00 \pm 0.00$	$5.00 \pm 0.01$	$5.00 \pm 0.01$	$5.00 \pm 0.00$	$5.00 \pm 0.01$	$5.00 \pm 0.01$	$5.00 \pm 0.00$	$5.00 \pm 0.01$	$5.02 \pm 0.01$
y	$4.00 \pm 0.00$	$4.00 \pm 0.01$	$4.02 \pm 0.01$	$4.00 \pm 0.00$	$4.00 \pm 0.01$	$4.01 \pm 0.01$	$4.00 \pm 0.00$	$4.01 \pm 0.01$	$4.01 \pm 0.01$
$n = 10$									
Noise [%]	0			25			50		
Uncertainty [%]	5	10	15	5	10	15	5	10	15
R	$2.01 \pm 0.00$	$2.01 \pm 0.00$	$2.01 \pm 0.01$	$2.00 \pm 0.00$	$2.02 \pm 0.00$	$2.03 \pm 0.01$	$2.01 \pm 0.00$	$2.03 \pm 0.00$	$2.04 \pm 0.01$
x	$5.00 \pm 0.00$	$5.00 \pm 0.00$	$5.00 \pm 0.00$	$5.00 \pm 0.00$	$5.01 \pm 0.01$	$4.99 \pm 0.01$	$5.00 \pm 0.00$	$5.00 \pm 0.01$	$4.99 \pm 0.01$
y	$4.00 \pm 0.00$	$4.01 \pm 0.00$	$4.00 \pm 0.01$	$4.00 \pm 0.00$	$4.00 \pm 0.00$	$4.00 \pm 0.01$	$4.00 \pm 0.00$	$4.00 \pm 0.01$	$4.00 \pm 0.01$
$n = 15$									
Noise [%]	0			25			50		
Uncertainty [%]	5	10	15	5	10	15	5	10	15
R	$2.01 \pm 0.00$	$2.01 \pm 0.00$	$2.01 \pm 0.00$	$2.01 \pm 0.02$	$2.01 \pm 0.00$	$2.02 \pm 0.00$	$2.01 \pm 0.00$	$2.02 \pm 0.00$	$2.02 \pm 0.00$
x	$5.00 \pm 0.00$	$5.00 \pm 0.00$	$5.00 \pm 0.01$	$5.00 \pm 0.00$	$5.00 \pm 0.00$	$5.00 \pm 0.01$	$5.00 \pm 0.00$	$5.01 \pm 0.00$	$4.99 \pm 0.01$
y	$4.00 \pm 0.00$	$4.00 \pm 0.00$	$4.01 \pm 0.01$	$4.00 \pm 0.00$	$4.00 \pm 0.00$	$4.00 \pm 0.01$	$4.00 \pm 0.00$	$4.00 \pm 0.00$	$4.00 \pm 0.01$
$n = 20$									
Noise [%]	0			25			50		
Uncertainty [%]	5	10	15	5	10	15	5	10	15
R	$2.01 \pm 0.01$	$2.01 \pm 0.00$	$2.02 \pm 0.00$	$2.01 \pm 0.01$	$2.01 \pm 0.00$	$2.03 \pm 0.00$	$2.02 \pm 0.02$	$2.02 \pm 0.00$	$2.03 \pm 0.00$
x	$5.00 \pm 0.00$	$5.00 \pm 0.00$	$4.99 \pm 0.00$	$5.00 \pm 0.00$	$5.00 \pm 0.00$	$5.00 \pm 0.00$	$5.00 \pm 0.00$	$5.00 \pm 0.00$	$5.00 \pm 0.01$
y	$4.00 \pm 0.01$	$4.00 \pm 0.00$	$4.00 \pm 0.01$	$4.01 \pm 0.00$	$4.00 \pm 0.00$	$4.00 \pm 0.01$	$4.00 \pm 0.00$	$4.00 \pm 0.00$	$4.00 \pm 0.01$

## 4 Conclusions

We examined the use of Gaussian sums for the reconstruction of a circle's center and radius, using its datapoints. The algorithm proposed proved to work well under all test scenarios (noise hits, inactive part of the detector, hit position uncertainties) and provide fast and accurate results.

## Acknowledgments

The research leading to these results has received funding from the FP7-INFRASTRUCTURES-2010, AIDA: Advanced European Infrastructures for Detectors at Accelerators, Grant Agreement No: 262025 and the European Union (European Social Fund ESF) and the Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) 2007-2013. ARISTEIA-1893-ATLAS MICROMEAS. The support is gratefully acknowledged.

## References

- [1] T. Altherr and J.Seixas, Cherenkov ring recognition using a non-adaptable network, Nuclear Instruments and Methods in Physics Research A317 (1992) 335-338
- [2] P. Volf et al., On mathematical approach to circle fitting problem in the analysis of rich detector data, Czech. J. Phys. 50/S1 (2000)
- [3] H. Babovsky, Recognition of ring pairs in the data analysis of Cherenkov detectors, Journal of Computational and Applied Mathematics 72 (1996) 227-234
- [4] T. Alexopoulos et al., Identification of circles from datapoints using the Legendre transform, Nuclear Instruments and Methods in Physics Research A745 (2014) 16-23